

## NRC Publications Archive Archives des publications du CNRC

### Enhancing the Outcome of Microarray-Based Molecular Diagnostics Tulpan, Dan; Léger, Serge

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. /  
La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version  
acceptée du manuscrit ou la version de l'éditeur.

#### **Publisher's version / Version de l'éditeur:**

*The Advances in Microarray Technology (AMT) [Proceedings], 2008*

**NRC Publications Archive Record / Notice des Archives des publications du CNRC :**  
<https://nrc-publications.canada.ca/eng/view/object/?id=8a5624cf-07a3-44eb-a42e-ef738c603bcf>  
<https://publications-cnrc.canada.ca/fra/voir/objet/?id=8a5624cf-07a3-44eb-a42e-ef738c603bcf>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at  
<https://nrc-publications.canada.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site  
<https://publications-cnrc.canada.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

**Questions?** Contact the NRC Publications Archive team at  
PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the  
first page of the publication for their contact information.

**Vous avez des questions?** Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la  
première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez  
pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.



National Research  
Council Canada

Conseil national  
de recherches Canada

Institute for  
Information Technology

Institut de technologie  
de l'information

# **NRC - CNRC**

---

## ***Enhancing the Outcome of Microarray-Based Molecular Diagnostics \****

Tulpan, D.C., Léger, S.  
May 2008

\* published at The Advances in Microarray Technology (AMT). May 7-8, 2008. Barcelona, Spain. NRC 50359. The Online Journal of Scientific Posters, 2008.

Copyright 2008 by  
National Research Council of Canada

Permission is granted to quote short excerpts and to reproduce figures and tables from this report, provided that the source of such material is fully acknowledged.

## Enhancing the Outcome of Microarray-Based Molecular Diagnostics

Dan Tulpan, Serge Léger

Knowledge Discovery Group, Institute of Information Technology, National Research Council of Canada, Moncton, New Brunswick, Canada.

### Abstract

Molecular diagnostics determines how genes and proteins are interacting in a cell and focuses upon gene and protein activity patterns in different types of cancerous or precancerous cells. Molecular diagnostics uncovers these sets of changes and captures this information as expression patterns via various data analysis techniques. The accuracy of the diagnostic process is hampered by many factors like: quality of the biological samples, accuracy of analytical methods and quality of microarray probes.

While the first two factors can be addressed by using better sample storage procedures and careful sample manipulation and by carefully choosing analytical techniques with wider domains of applicability and proven higher accuracies, the third factor is most of the time hard to adjust, given the rigidity of existing microarray platforms and increased costs for their re-design and re-validation.

Here, it is presented a technique that allows researchers to use experimental data (expression values) obtained with existing microarray platforms whose probe sets may not match the biological information known today. The technique is capable of adjusting the expression values in the experimental data by spotting inconsistencies between probes and their sequence origins and estimating their level of hybridization using molecular thermodynamic modeling.

### Summary

Molecular diagnostics uncovers gene and protein activity patterns in different types of cancerous cells via data analysis techniques. The accuracy of the diagnostic process can be improved by adjusting the expression values in the experimental data using molecular thermodynamic modeling.

### Speaker Biography

Dan Tulpan is a Research Officer at the Department of Information Technology of the National Research Council of Canada. He received a Ph.D. from the Computer Science Department of the University of British Columbia (Canada) in 2006 and was a Postdoctoral Fellow in Brinkman Laboratory at Simon Fraser University and instructor for the Masters of Software Systems Program at UBC. Dan's research interests are currently focused on topics of bioinformatics, biotechnology, artificial intelligence and empirical algorithms.

# Enhancing the Outcome of Microarray-Based Molecular Diagnostics

Dan Tulpan & Serge Léger  
 Knowledge Discovery Group, Institute of Information Technology, National Research Council of Canada, Moncton, New Brunswick, Canada

## Background

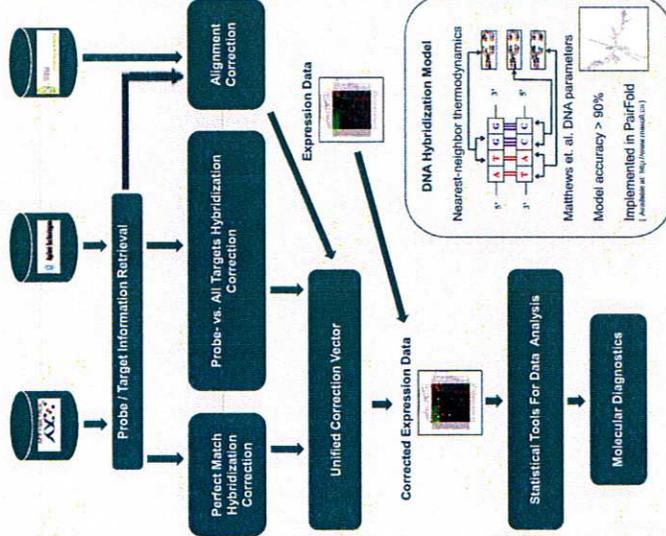
Molecular diagnostics determines how genes and proteins are interacting in a cell and focuses upon gene and protein activity patterns in different types of cancerous or precancerous cells. Molecular diagnostics uncovers these sets of changes and captures this information as expression patterns via various data analysis techniques. The accuracy of the diagnostic process is hampered by many factors like: quality of the biological samples, accuracy of analytical methods and quality of microarray probes.

While the first two factors can be dealt with by using better sample storage procedures and careful sample manipulation and by carefully choosing analytical techniques with wider domains of applicability and proven higher accuracies, the third factor is most of the time hard to adjust, given the rigidity of existing microarray platforms and increased costs for their redesign and revalidation.

Here, it is presented a technique that allows researchers to use experimental data (expression values) obtained with existing microarray platforms whose probe sets may not match the biological information known today. The technique is capable of adjusting the expression values in the experimental data by spotting inconsistencies between probes and their sequence origins and estimating their level of hybridization using molecular thermodynamic models.

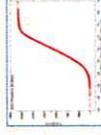
## Methods

The following computational pipeline is proposed to adjust expression values based on platform dependent probe/target information and publicly available knowledge.



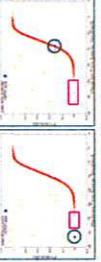
## Results

**Perfect Match Hybridization Correction**  
 Perfect match stabilities vary for different platforms, thus leading to different expression values and different sensing capabilities within the same platform.



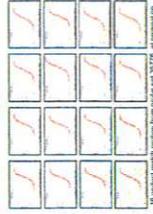
Alignment Chip: HG\_U10642

**Probe vs. All Targets Hybridization Correction**  
 Perfect match probes from different probe sets do not capture the correct target.



A "wild" and a "real" probe (numbers are 50.7% of present in Affymetrix Chip: HG\_U10642)

For most perfect match probes there is a set of targets that can potentially hybridize with the probe, thus competing with the desired target (highlighted in purple).



Alignment Chip: HG\_U10642 (2 "wild" probes)

### Alignment Correction

A simple sequence similarity check reveals a large number of one-to-many relationships between target sequences used in probe design and currently known sequence information.



## Future Directions

A thorough evaluation, comparison and in vivo validation of the current approach is under investigation.

We are open for collaborations with academic and industrial partners. If you are interested in this technology please contact: Dr. Dan Tulpan at [dan.tulpan@nrc-cnrc.gc.ca](mailto:dan.tulpan@nrc-cnrc.gc.ca)